sciendo
Sciendo is a De Gruyter company

Contributions to Tobacco & Nicotine Research
formerly: Beiträge zur Tabakforschung International
Volume 34 · No. 1 · January 2025
DOI: 10.2478/cttr-2025-0003

A Statistical Approach for Comparative Assessment of the Effect of Smoke Exposure in *In Vivo* Experiments: A Case Study of an OECD 90-Day Inhalation Study Including 3R4F and 1R6F Reference Cigarettes*

by

Athanasios Kondylis[1], *Ulrike Kogel*[1], *Jenny Ho*[2], *EeTsin Wong*[2], *Blaine Phillips*[2], *Julia Hoeng*[1], *and Patrick Vanscheeuwijck*[1]

[1] *PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland*
[2] *Vectura Fertin Pharma Laboratories Pte. Ltd., Science Park II, Singapore*

SUMMARY

In vivo testing is a crucial part of product development. Most *in vivo* toxicology studies are conducted in a heavily regulated environment, in accordance with sound scientific principles, and in an ethical manner. Statistical methods are an important component of these scientific principles, and many regulatory authorities provide guidelines or recommendations for statistical analyses and reporting. We propose a rigorous and comprehensive statistical approach to assess the impact of smoke exposure during *in vivo* comparative studies using statistical equivalence with variable equivalence limits based on historical data. The analysis enables comparison of the effect of a new/test aerosol to its reference/control while leveraging information about reference item variation across multiple studies. This approach is helpful in assessing the relevance of observed effects while incorporating sources of variability estimated using data from the current and previous experiments. The proposed method was used to analyze an OECD (Organization for Economic Co-operation and Development) 90-day inhalation study to determine the effects of exposure to smoke emitted by the new 1R6F reference cigarette as compared to the older 3R4F reference cigarette. Data from previous OECD 90-day *in vivo* studies using 3R4F cigarette smoke were used to improve our assessment of the effects observed in the current study. [Contrib. Tob. Nicotine Res. 34 (2025) 26–33]

KEYWORDS

In vivo testing; reference cigarette smoke; statistical equivalence; statistical intervals

ZUSAMMENFASSUNG

In vivo Experimente sind ein wichtiger Bestandteil der Produktentwicklung. Toxikologische *in vivo* Studien unterliegen, in Übereinstimmung mit wissenschaftlichen und ethischen Prinzipien, strengen regulatorischen Auflagen. Ein wichtiger Teil dieser Prinzipien ist die richtige Anwendung statistischer Analysen, welche oft von den entsprechenden zuständigen regulatorischen Behörden als Leitfäden oder Empfehlungen zur Verfügung gestellt werden.

*Received: 5th July 2024 – accepted: 8th November 2024

Um die Auswirkung von Rauch in *in vivo* Vergleichsstudien zu bewerten, schlagen wir die Verwendung des statistischen Äquivalenztests mit variablen Äquivalenzgrenzen basierend auf Vergangenheitsdaten als präzisen und umfassenden statistischen Ansatz vor.
Diese Analyse erlaubt den Vergleich von neuen oder Test-Aerosolen mit den Referenzendaten und deren Abweichungsdaten, welche in vorherigen Studien gesammelt wurden. Auf diese Weise kann die Relevanz von beobachteten Effekten besser bewertet werden, da Variationsquellen der jetzigen Studie als auch von vorhergegangenen Studien miteinbezogen werden.
Zur Veranschaulichung haben wir diese Methode bei der Auswertung einer OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)-konformen 90-Tage Inhalationstudie angewendet, um die Effekte, die bei Rauchexponierung durch die neue Referenzzigarette 1R6F auftreten, mit den Effekten der vorhergehenden Referenzzigarette 3R4F zu vergleichen. Daten von früheren OECD-konformen 90-Tage Inhalationstudien, welche mit der Referenzzigarette 3R4F durchgeführt worden waren, wurden dazu verwendet, die Effekte in der neuen Studie besser auswerten zu können. [Contrib. Tob. Nicotine Res. 34 (2025) 26–33]

RESUME

L'évaluation scientifique des données des études *in vivo* constituent un élément crucial du développement de produits. Les études toxicologiques *in vivo* sont généralement menées dans un environnement fortement réglementé, conformément à des principes scientifiques solides et de manière éthique. Les méthodes statistiques constituent un élément essentiel de ces principes scientifiques, et des lignes directrices ou des recommandations pour les analyses et rapports statistiques sont souvent fournies par les autorités de réglementation. Nous proposons une approche statistique rigoureuse et complète pour évaluer l'impact de l'exposition à la fumée dans des études comparatives *in vivo* utilisant l'équivalence statistique avec des limites d'équivalence variables basées sur des données historiques. L'analyse permet de comparer l'effet d'un aérosol nouveau/test à sa référence/contrôle tout en exploitant les informations sur la variation de l'élément de référence dans plusieurs études. De cette manière, nous pouvons mieux évaluer la pertinence des effets observés tout en intégrant les sources de variabilité estimées à partir des données des expériences actuelles et précédentes. Pour illustrer la méthode proposée, nous l'avons utilisée pour analyser une étude d'inhalation de 90 jours afin de déterminer les effets de l'exposition à la fumée émise par la nouvelle cigarette de référence 1R6F par rapport à l'ancienne cigarette de référence 3R4F. Les données d'études *in vivo* précédentes de 90 jours de l'OCDE (Organisation de coopération et de développement économiques) utilisant la fumée de cigarette 3R4F ont été utilisées pour améliorer notre évaluation des effets observés dans la présente étude. [Contrib. Tob. Nicotine Res. 34 (2025) 26–33]

INTRODUCTION

In vivo testing is a critical step in the assessment of the toxicological effects of experimental test aerosols, including cigarette smoke. Existing regulatory guidelines define good practices and requirements for *in vivo* inhalation studies to ensure they are conducted in a scientifically sound and humane manner. The Organization for Economic Co-operation and Development (OECD) Test Guideline (TG) 413 was designed to fully characterize test article toxicity when administered by the inhalation route for 90 days and to provide robust data for inhalation risk assessments (1).
Reference cigarettes from the University of Kentucky are widely used for the toxicological assessment of the effects of cigarette smoke, with the 3R4F reference cigarette serving as the general standard for several decades. Recently, the 1R6F was developed and manufactured to replace the 3R4F. Here, we propose a novel method for comparative *in vivo* toxicological assessment of cigarettes or novel tobacco products. We used *in vivo* toxicology data following repeated dose inhalation exposure to the reference cigarette 3R4F and the new reference cigarette 1R6F and compared a large number of biomarkers and endpoints relating to: test atmosphere analysis, respiratory physiology, exposure endpoints, urinary nicotine metabolites, inflammatory cells, and clinical chemistry. While standard statistical methodologies only use current study data to test for differences, the proposed methodology performs equivalence testing and employs historical 3R4F data to set up equivalence ranges and limits.
The current study data enable direct comparison of effects between the two cigarettes including average differences and study-specific variability estimates; historical 3R4F data provide estimates for the historical or long-term variability of the reference cigarette. We leveraged the historical data to improve our interpretation of the observed differences between the two cigarettes in the current study. Including historical data is reasonable because even under a statistically sound design (e.g., block randomized designs), data variation cannot be fully attributed to pre-identified study design parameters. Variation inherent to smoke and aerosol generation emanates from multiple and often unknown sources during aerosol generation or the bioanalytical quantification process. Cigarette smoke is a highly complex system, with variation that results from many uncontrollable sources (2–4). In addition, despite all efforts to optimize and validate analytical methods and instrumentation, many approaches still provide measurements with non-negligible error. Although increasing the sample size could minimize variation while performing traditional statistical comparisons, this is not compatible with animal welfare concerns. The use of historical data provides a powerful and ethically acceptable alternative.
Clear regulatory directives on setting equivalence limits for comparative tobacco product assessment do not exist. Indeed, there are no widely accepted statistical methodologies that could be used to demonstrate equivalence for tobacco products (5). Our proposal consists of following the equivalence principle for statistical comparisons while

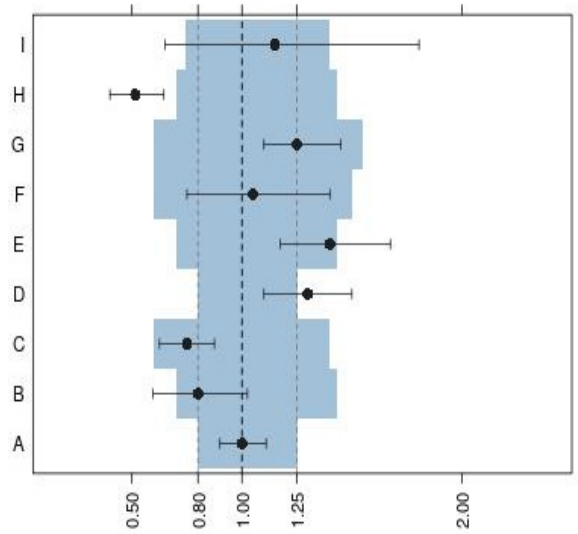

Figure 1. Illustrated cases of the relative effect of a test item *versus* a reference item as provided by their geometric mean ratio (black circles) and associated 90% confidence intervals (black bars). The blue boxes show the variability of the reference item. Rows A to I illustrate different scenarios (explained further in the main text) for the value of the mean ratio and confidence interval and how they influence conclusions about the differences and equivalence of the two items.

expanding the range beyond the standard bioequivalence limits to incorporate the long-term variability of the reference product. The use of variable equivalence limits based on historical data evidence has been discussed in the pharmaceutical industry (6–8). Variable equivalence limits have been also proposed for comparisons of smoke constituent yields and *in vitro* comparative assessment (2, 9).

METHODS

Statistical approach

The term "comparative assessment" refers to testing for differences, which in statistical terms translates to testing the null hypothesis that products do not differ *versus* the alternative hypothesis of product difference. Detection of statistically significant differences leads to rejection of the null hypothesis. This is commonly achieved using probability evidence provided by the so-called (*p*)-values. For a significance level (α), one declares differences to be statistically significant when the *p*-value is smaller than α. The level α is also known as the consumer's risk or the probability of falsely concluding that a difference exists when there is none. As the number of out-comes on which the products are compared increases, so does the consumer's risk, exceeding the level α, so the *p*-values must be adjusted (10). The use of *p*-values is contentious, and the literature includes extensive discussions on their use and misuse (11–14). In a contrasting approach, equivalence testing has been used to demonstrate equivalence between two products (15, 16). This approach has mainly been applied in absorption studies during drug development under the name of bioequivalence, in accordance with U.S. Food and Drug Administration (FDA) and European regulatory authority guidelines (17–20). The starting point (null hypothesis) in equivalence testing is that the two products are different, so they are declared equivalent with rejection of the null hypothesis. Two one-sided tests (TOST) (21, 22) are used for this purpose, in contrast to the single two-sided test performed in difference testing.

Confidence intervals are used to assess differences and/or equivalence (23, 24), providing more insight into the comparative assessment between two products. They provide information on the direction and the magnitude of the observed differences, as well as the associated uncertainty, which is reflected in their width. The width of the confidence interval depends on the sample size and confidence level 1−α, creating a direct link between confidence intervals and statistical tests. For difference testing, we look for a confidence interval on the (1−α)% confidence level; for testing equivalence with TOST we look for a (1−2α)% confidence interval. Figure 1 illustrates several hypothetical confidence intervals and their link to statistical difference and equivalence testing resulting from a comparison of mean ratios between two products. Row A indicates no difference between the two products with high certainty. The mean ratio between the two products is equal to one, and the confidence interval is narrow. The confidence interval in row H is equally narrow, but the effect of the test product is reduced by almost 50% relative to the reference product. In terms of a difference testing approach, the observed effect in case H is highly significant, while the one in case A is not. Rows C and G illustrate statistically significant changes; a reduction and an increase for rows C and G, respectively. Statistically significant increase is also depicted in row D and row E. For equivalence testing, case H shows no equivalence between the two products, whereas case A shows equivalence within the standard equivalence limits of [0.8, 1.25]. No equivalence is the conclusion for all rows other than A, given all confidence intervals don't entirely lie within the equivalence zone of [0.8, 1.25].These equivalence limits are commonly used in bioequivalence studies and are also recommended by the FDA. Note that they are symmetric in the ratio scale (1/0.8 = 1.25). We propose to extend the equivalence limits beyond [0.8, 1.25] according to:

$$\left[\delta_L, \delta_U\right] = \exp\left\{\pm\left(\log(1.25), \left(t_{\alpha,df} + t_{\beta/2,df}\right)\frac{\sigma_R}{\sqrt{n}}\right)\right\} \quad [1]$$

where σ_R represents long-term variability (in standard deviation terms), the constants t_α and $t_{\beta/2}$ represent the student percentiles at consumer's and producer's risks α and β, respectively, with $df = 2n-2$ degrees of freedom, where $2n$ is the total sample of animals allocated to the two products. For illustration purposes we report in Table 1 and Table 2 variable lower and upper equivalence limits using two different proposals (using equation [1] and an additional approach described in the Supplementary Material) under different statistical design parameters. The Supplementary Material provides a more detailed description, references, and insights into the equation above and the statistical methods used within the proposed framework. In Figure 1, this approach is illustrated by overlaying blue boxes to indicate the new (variable) equivalence range

Table 1. Lower and upper equivalence range limits (δ_L and δ_U, respectively) for varying statistical design parameters: sample size by group (n), consumer's risk (α), producer's risk (β), and coefficient of variation (cv) with corresponding standard deviation (σ_R).

n	α	β	cv	σ_R	δ_L	δ_U
10	0.05	0.05	0.2	0.198	0.786	1.271
10	0.05	0.05	0.3	0.294	0.700	1.428
10	0.05	0.05	0.4	0.385	0.627	1.596
10	0.05	0.05	0.5	0.472	0.564	1.773
10	0.05	0.1	0.2	0.198	0.805	1.243
10	0.05	0.1	0.3	0.294	0.725	1.380
10	0.05	0.1	0.4	0.385	0.655	1.526
10	0.05	0.1	0.5	0.472	0.596	1.679

Table 2. Lower and upper alternative equivalence range limits (δ_L^* and δ_U^*, respectively) for varying statistical design parameters: constant k and coefficient of variation (cv) with corresponding standard deviation (σ_R).

k	cv	σ_R	δ_L^*	δ_U^*
1	0.2	0.198	0.820	1.219
1	0.3	0.294	0.746	1.341
1	0.4	0.385	0.680	1.470
1	0.5	0.472	0.624	1.604
0.76	0.2	0.198	0.860	1.162
0.76	0.3	0.294	0.800	1.250
0.76	0.4	0.385	0.746	1.340
0.76	0.5	0.472	0.698	1.432

limits. Comparing the confidence intervals and blue boxes in Figure 1, the interpretation of the statistical comparisons' changes concern rows C, F, and G. In all three cases, the two products would be considered equivalent with respect to the new (variable) equivalence range limits, while using the [0.8,1.25] equivalence limits they would not. Rows F and I illustrate inconclusive cases where neither statistical difference nor statistical equivalence is met, when the standard equivalence limits are used. Yet, given the new (variable) equivalence range limits, row F would classify as equivalent.

Data from a 90-day OECD rat inhalation study served as a use case for our statistical approach on the comparative assessment of the exposure effect to smoke from the standard reference cigarette 3R4F *versus* the new reference cigarette 1R6F. The study data and analytical methods used for their generation are described below. In addition to the current study data, historical 3R4F data on the same set of endpoints were extracted from three previous inhalation studies and were gathered to build the historical reference data set. Starting from the current study data, the geometric mean ratio between the new reference cigarette over the standard one was computed for each endpoint (Supplementary Material I). The statistical comparisons were performed on the ratio scale, and 90% confidence intervals for the geometric mean ratio were computed for equivalence testing. Equivalence is met when the entire confidence interval on the geometric mean ratio for an endpoint lies within the equivalence limits. These are defined using the historical 3R4F data and expand the equivalence limits beyond the [0.8, 1.25] range using Equation 1. For illustrative purposes, the long-term variability (σ_R) estimates for all the analyzed endpoints on 3R4F (estimated across the three previous inhalation studies) are plotted in Figure 2.

Case data generation

The analysis presented in this work focuses on the comparison of two products: new (1R6F) *versus* reference (3R4F) in a 90-day OECD rat inhalation study that followed OECD test guideline 413 (1) and was conducted in compliance with the OECD Principles of Good Laboratory Practice and the test facility's quality management system. The test facility is National Parks Board/Animal and Veterinary Service (NParks/AVS-licensed) and Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) accredited, and care and use of the rats was in accordance with the National Advisory Committee for Laboratory Animal Research (NACLAR) Guideline (25) and AAALAC requirements (26). All animal experiments were approved by the Institutional Animal Care and Use Committee.

The cigarette smoke was administered by nose-only exposure inhalation to outbred male and female Sprague Dawley rats [Crl:CD(SD)] bred under specific pathogen-free conditions (10 rats per sex). The exposure was conducted for approximately 13 weeks (6 hours per day, 5 days per week). Week 1 consisted of a time-adaptation phase during which the rats were exposed to increasing exposure durations over 7 days. All exposures were targeted to deliver 23 µg/L nicotine in the test atmosphere.

A set of relevant endpoints was analyzed as part of the comparative assessment between 3R4F and 1R6F. These endpoints were grouped into four main categories:

(i) test atmosphere characteristics in the exposure chambers;
(ii) cigarette smoke uptake parameters as respiratory physiology, carboxyhemoglobin levels in the blood (as a marker for CO uptake), and urinary nicotine metabolite levels;
(iii) inflammatory biomarkers as inflammatory cells;
(iv) clinical chemistry markers.

Data from selected endpoints from three historical studies that used 3R4F (27–29) were also collected and analyzed as part of the comparative assessment between 3R4F and 1R6F.

A tabulated method description for the analyzed parameters is given in the Supplementary Material II (Supplemental Tables A–C).

RESULTS

The results of the statistical comparisons from the proposed comparative assessment are shown in Figures 3 and 4. The confidence intervals depict the estimates and the uncertainty related to the observed differences between the two reference cigarettes, while the blue boxes reflect equivalence zones for each endpoint derived from the historical data variability of the historical (3R4F) reference product. Figure 3 provides strong evidence for equivalence between the 3R4F and 1R6F cigarettes for all test atmosphere endpoints except formaldehyde. The observed differences

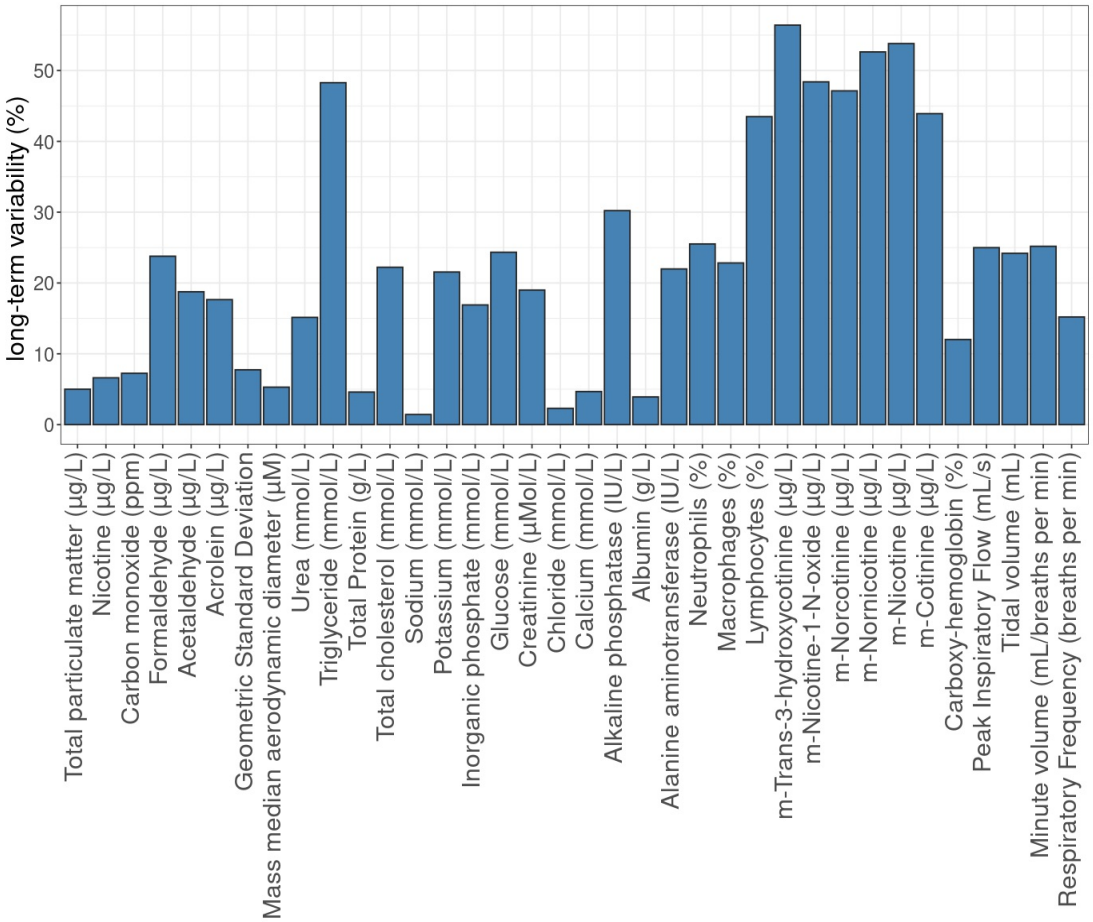

Figure 2. Long-term variability estimates (σ_R) depicted in blue bars for all endpoints across the historical 3R4F data. Variability estimates are expressed in percentages.

are estimated with high precision for these endpoints, as reflected by the relatively narrow confidence intervals. The only exceptions in this case are the carbonyl compounds, especially formaldehyde, which was also significantly increased in 1R6F as compared to 3R4F (9, 30). The blue boxes are fixed at the standard limits of [0.8, 1.25], reflecting that no extra variability is present in the 3R4F reference cigarette results across the historical data. Figure 4 highlights the diversity of the obtained results for the various biomarker groups. Variability is observed in both the current study data (as reflected by the large confidence intervals) and the historical data (as reflected by the large equivalence ranges depicted as blue boxes). Figure 4 highlights that for all the biomarkers, the average observed differences between 1R6F and 3R4F are within the equivalence ranges and, therefore, within the variability range of the reference product. These results can be visually confirmed by noting that the geometric mean ratio estimates (black circles) fall within the equivalence limits (blue boxes). However, the lower and/or upper confidence limits are not always within the equivalence ranges. In these cases, equivalence, as mathematically defined, is not met. In our case study equivalence is confirmed for all respiratory physiology and exposure endpoints. Yet, it is not always met for nicotine metabolites, inflammatory cells, and clinical chemistry endpoints. For all studied

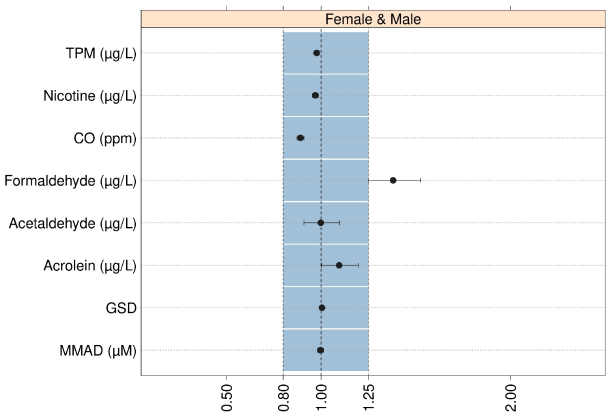

Figure 3. Geometric mean ratio estimates (black circles) and their associated 90% confidence intervals for smoke emissions from the 1R6F reference cigarette over the 3R4F reference cigarette for test atmosphere endpoints. Blue boxes define equivalence ranges for the geometric mean ratio estimate based on the variability of 3R4F estimated using historical data. The standard equivalence limits of 0.8 and 1.25 and the reference value of 1 are shown (black dashed lines). TPM: Total Particulate Matter, CO: Carbon monoxide, GSD: Geometric Standard Deviation and MMAD: Mass Median Aerodynamic Diameter.

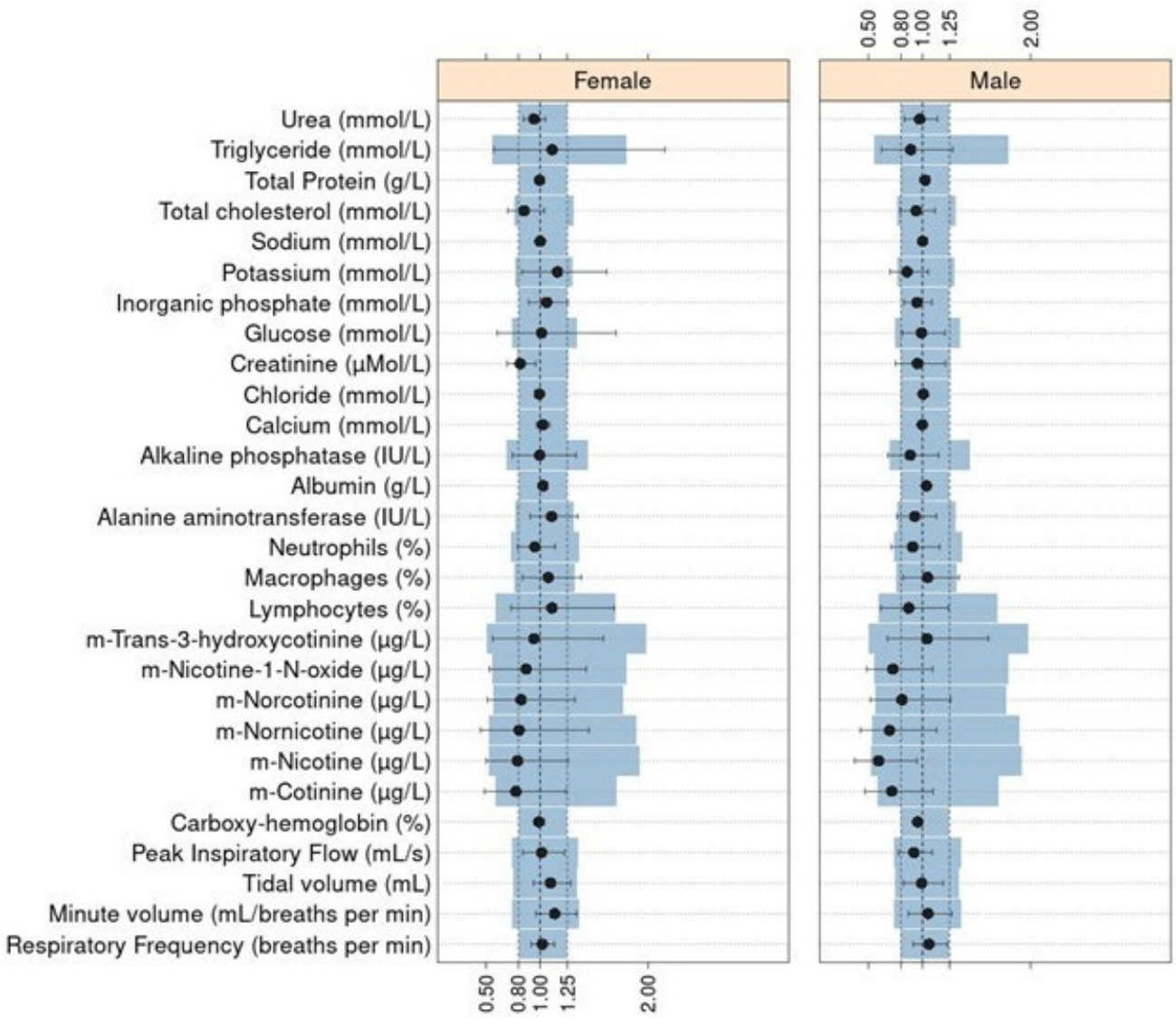

Figure 4. Geometric mean ratio estimates (black circles) and their associated 90% confidence intervals for smoke from the 1R6F reference cigarette over the 3R4F reference cigarette for female (left panel) and male (right panel) rats across all endpoints related to (from bottom to top): respiratory physiology, exposure, urinary nicotine metabolites, and clinical chemistry. Blue boxes define equivalence ranges for the geometric mean ratio as estimated based on the variability of 3R4F estimated using historical data. The standard equivalence limits of 0.8 and 1.25 and the reference value of 1 are shown (black dashed lines).

endpoint categories, with the exception of the urinary nicotine metabolites, there is no evidence of an increasing or decreasing trend. Nicotine metabolites show a consistent inferiority trend.

DISCUSSION

This manuscript describes a novel statistical comparison approach for tobacco product comparative assessment using *in vivo* exposure studies. The approach was illustrated using data from an OECD 90-day inhalation study using two reference cigarettes. The proposed methodology allowed a comparative assessment between 3R4F and 1R6F reference cigarettes using the equivalence principle with variable equivalence limits based on historical data variation. This is a rigorous way to assess product differences when historical data are available. The proposed method combined the study data from the OECD 90-day inhalation study together with the historical data on the 3R4F reference product to better interpret the current studyresults and improve the product comparisons. Properly setting the equivalence limits and selecting the appropriate way to integrate all prior information about the reference product remain critical points for further investigation.

The OECD study results show that differences observed between the two reference cigarettes were modest and on average within the range of the variability of 3R4F. The analysis of the *in vivo* study provided substantial scientific evidence that the 1R6F reference cigarette is a suitable replacement for the 3R4F reference cigarette in comparative tobacco product assessments. However, the statistical analysis results did not provide the necessary mathematical evidence for proving formal equivalence between the two cigarettes across all biomarkers under investigation.

CONCLUSIONS

Equivalence analyses with variable equivalence limits using historical data may largely improve the comparative assessment of smoke/aerosol exposure in *in vivo* comparative studies. It combines current study evidence, as provided by the study data, and uses historical data on the reference product to scale the observed findings and better assess the relevance of the observed effects. This was demonstrated with the analysis of an OECD 90-day inhalation study using the 3R4F and 1R6F reference cigarettes.

SUPPLEMENTARY INFORMATION

This publication contains supplementary material available at https://sciendo.com/article/10.2478/cttr-2025-0003.

ACKNOWLEDGMENTS

The authors would like to thank the two anonymous referees, as well as Irfan Gunduz and Maxim Belushkin for their very helpful suggestions. We would also like to thank the Publications Team in PMI for their precious help in editing this manuscript.

FINANCIAL SUPPORT

The research described in this manuscript was sponsored by Philip Morris Products S.A.

CONFLICTS OF INTEREST

All authors were employed by Philip Morris International or Vectura Fertin Pharma Laboratories (a subsidiary of Philip Morris International) at the time of this work.

REFERENCES

1. Organisation for Economic Co-operation and Development (OECD): Test No. 413: Subchronic Inhalation Toxicity: 90-Day Study; OECD Guidelines for the Testing of Chemicals, Section 4, OECD, Paris, France, 2018. DOI: 10.1787/9789264070806-en
2. Belushkin, M., G. Jaccard, and A. Kondylis: Considerations for Comparative Tobacco Product Assessments Based on Smoke Constituent Yields; Regul. Toxicol. Pharmacol. 73 (2015) 105–113. DOI: 10.1016/j.yrtph.2015.06.017
3. Eldridge, A., T.R. Betson, M.V. Gama, and K. McAdam: Variation in Tobacco and Mainstream Smoke Toxicant Yields from Selected Commercial Cigarette Products; Regul. Toxicol. Pharmacol. 71 (2015) 409–427. DOI: 10.1016/j.yrtph.2015.01.006
4. Verron, T., X. Cahours, and S. Colard: Extension of Critical Difference for Product Comparison. Application to Tobacco Products; Beitr. Tabakforsch. Int.. 28 (2019) 310–316. DOI: 10.2478/cttr-2019-0012
5. U.S. Food and Drug Administration (FDA), Center for Tobacco Products (CTP): Guidance for Industry and FDA Staff, Section 905(j) Reports: Demonstrating Substantial Equivalence for Tobacco Products; FDA, Silverspring, MD, USA, 2011, 14 pp. Available at: https://www.federalregister.gov/documents/2011/01/06/2011-35/guidance-for-industry-and-food-and-drug-administration-staff-section-905j-reports-demonstrating
6. Dangi, Y.S., M.L. Soni, and K.P. Namdeo: Highly Variable Drugs: Bioequivalence Requirements and Regulatory Perspectives; Int. J. Curr. Pharm. Res. 3 (2010) 24–28.
7. Endrenyi, L. and L. Tothfalusi: Bioequivalence for Highly Variable Drugs: Regulatory Agreements, Disagreements, and Harmonization; J. Pharmacokinet. Pharmacodyn. 46 (2019) 117–126. DOI: 10.1007/s10928-019-09623-w
8. Endrenyi, L. and L. Tothfalusi: Regulatory and Study Conditions for the Determination of Bioequivalence of Highly Variable Drugs; J. Pharm. Pharm. Sci. 12 (2009) 138–149. DOI: 10.18433/j3zw2c
9. Sakai, Y., S. Mori, M. Yanagimachi, T. Takahashi, K. Shibuya, A. Kumagai, S. Ishikawa, S. Ito, and T. Fukushima: Inter-Laboratory Reproducibility and Interchangeability of 3R4F and 1R6F Reference Cigarettes in Mainstream Smoke Chemical Analysis and *In Vitro* Toxicity Assays; Contrib. Tob. Nicotine Res. 29 (2020) 119–135. DOI: 10.2478/cttr-2020-0011
10. Miller, R.G.: Simultaneous Statistical Inference; Springer, New York, NY, USA, 1981, 299 pp., ISBN: 9780387905488
11. Kim, J. and H. Bang: Three Common Misuses of *P* Values; Dent. Hypotheses 7 (2016) 73–80. DOI: 10.4103/2155-8213.190481
12. Brereton, R.G.: The Use and Misuse of P Values and Related Concepts; Chemometr. Intell. Lab. Syst. 195 (2019) 103884. DOI: 10.1016/j.chemolab.2019.103884
13. Schervish, M.J.: *P* values: What They Are and What They Are Not; Am. Stat. 50 (1996) 203–206. DOI: 10.2307/2684655
14. Wasserstein, R.L. and N. A. Lazar: The ASA Statement on *p*-Values: Context, Process, and Purpose; Am. Stat. 70 (2016) 129–133. DOI: 10.1080/00031305.2016.1154108
15. Anderson-Cook, C.M. and C.M. Borror: The Difference between "Equivalent" and "Not Different"; Qual. Eng. 28 (2016) 249–262. DOI: 10.1080/08982112.2015.1079918
16. Barros, J.A.O.: Report on Indirect Method to Obtain Stress-Strain Response of Fiber-Reinforced Concrete (FRC); American Concrete Institute (ACI), Farmington Hills, MI, USA, 2016, 4 pp.
17. U.S. Food and Drug Administration (FDA): Guidance for Industry on Statistical Approaches to Establishing Bioequivalence, Availability; FDA, Silver Spring, MD, USA, 2001, Federal Register, 66, No.23, 8805–8806.
18. Chow, S.C. and J.P. Liu: Recent Statistical Developments in Bioequivalence Trials - A Review of the FDA Guidance; Drug Inf. J. 28 (1994) 851–864. DOI: 10.1177/009286159402800321

19. Guimarães Morais, J.A. and M. do Rosário Lobato: The New European Medicines Agency Guideline on the Investigation of Bioequivalence; Basic Clin. Pharmacol. Toxicol. 106 (2010) 221–225. DOI: 10.1111/j.1742-7843.2009.00518.x
20. European Union: Directive 2014/40/EU of the European Parliament and of The Council of 3 April 2014 on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning the Manufacture, Presentation and Sale of Tobacco and Related Products and Repealing Directive 2001/37/EC; Off. J. Eur. Union (2014) L127/1–L127/38.
21. Schuirmann, D.J.: A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability; J. Pharmacokinet. Biopharm. 15 (1987) 657–680. DOI: 10.1007/BF01068419
22. Chow, S.-C. and J.-P. Liu: Design and Analysis of Bioavailability and Bioequivalence Studies; 1st edition, Marcel Dekker, New York, NY, USA, 1992, 416 pp. ISBN: 9780824786823
23. Berger, R.L. and J.C. Hsu: Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets; Stat. Sci. 11 (1996) 283–319. DOI: 10.1214/ss/1032280304
24. Hsu, J.C., J.T. Gene Hwang, H.K. Liu, and S.J. Ruberg: Confidence Intervals Associated with Tests for Bioequivalence; Biometrika 81 (1994) 103–114. DOI: 10.2307/2337054
25. National Advisory Committee for Laboratory Animal Research (NACLAR): NACLAR Guidelines. 2004; Available at: https://www.nparks.gov.sg/avs/animals/animals-in-scientific-research/naclar-guidelines/naclar-guidelines (accessed November 2024).
26. National Research Council 2011. Guide for the Care and Use of Laboratory Animals: 8th edition, The National Academies Press, Washington, DC, USA, 220 pp. DOI: 10.17226/12910
27. Wong, E.T., U. Kogel, E. Veljkovic, F. Martin, Y. Xiang, S. Boue, G. Vuillaume, P. Leroy, E. Guedj, G. Rodrigo, N.V. Ivanov, J. Hoeng, M.C. Peitsch, and P. Vanscheeuwijck: Evaluation of the Tobacco Heating System 2.2. Part 4: 90-Day OECD 413 Rat Inhalation Study with Systems Toxicology Endpoints Demonstrates Reduced Exposure Effects Compared with Cigarette Smoke; Regul. Toxicol. Pharmacol. 81 Suppl 2 (2016) S59–S81. DOI: 10.1016/j.yrtph.2016.10.015
28. Oviedo, A., S. Lebrun, U. Kogel, J. Ho, W.T. Tan, B. Titz, P. Leroy, G. Vuillaume, M. Bera, F. Martin, G. Rodrigo, M. Esposito, R. Dempsey, N.V. Ivanov, J. Hoeng, M.C. Peitsch, and P. Vanscheeuwijck: Evaluation of the Tobacco Heating System 2.2. Part 6: 90-Day OECD 413 Rat Inhalation Study with Systems Toxicology Endpoints Demonstrates Reduced Exposure Effects of a Mentholated Version Compared with Mentholated and Non-mentholated Cigarette Smoke; Regul. Toxicol. Pharmacol. 81 Suppl. 2 (2016) S93–S122. DOI: 10.1016/j.yrtph.2016.11.004
29. Phillips, B.W., W.K. Schlage, B. Titz, U. Kogel, D. Sciuscio, F. Martin, P. Leroy, G. Vuillaume, S. Krishnan, T. Lee, E. Veljkovic, A. Elamin, C. Merg, N.V. Ivanov, M.C. Peitsch, J. Hoeng, and P. Vanscheeuwijck: A 90-Day OECD TG 413 Rat Inhalation Study with Systems Toxicology Endpoints Demonstrates Reduced Exposure Effects of the Aerosol from the Carbon Heated Tobacco Product version 1.2 (CHTP1.2) Compared with Cigarette Smoke. I. Inhalation Exposure, Clinical Pathology and Histopathology; Food Chem. Toxicol. 116 (2018) 388–413. DOI: 10.1016/j.fct.2018.04.015
30. Jaccard, G., D.T. Djoko, A. Korneliou, R. Stabbert, M. Belushkin, and M. Esposito: Mainstream Smoke Constituents and *In Vitro* Toxicity Comparative Analysis of 3R4F and 1R6F Reference Cigarettes; Toxicol. Rep. 6 (2019) 222–231. DOI: 10.1016/j.toxrep.2019.02.009

Corresponding author:

Athanasios Kondylis
PMI R&D
Philip Morris Products S.A.
Quai Jeanrenaud 5
2000 Neuchâtel
Switzerland
E-mail: athanasios.kondylis@pmi.com